GMR
The Original

TLR4 AND PTGS2 AS A SHARED BLOOD TRANSCRIPTOMIC SIGNATURE LINKING TYPE 2 DIABETES MELLITUS AND ATHEROSCLEROSIS: AN INTEGRATED ANALYSIS OF INDEPENDENT GEO DATASETS

Aparna Kalyanaraman[1], Rajeswary Hari[1*], Sri Devi Masilamani[1], Kowsalya Rajendran[2]

[1]Department of Biotechnology, Dr. M. G. R. Educational and Research Institute, Maduravoyal, Chennai –600095. Tamil Nadu, India
[2] Centre for Applied Research, Saveetha school of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Thandalam, Chennai- 602105, Tamil Nadu, India
*Corresponding email – rajihar@gmail.com

ABSTRACT

Background: Type 2 diabetes mellitus (T2DM) and atherosclerosis (AS) share common inflammatory and metabolic pathways, and common blood-based transcriptomic signals may help with earlier and more integrated cardiometabolic risk assessment. We evaluated whether such a signal can be detected using real, independent, publicly deposited blood transcriptomic datasets, and report the result plainly, including where the evidence is strong and where it is modest.

Methods: Two independent datasets were obtained from Gene Expression Omnibus (GEO): GSE95849 (T2DM patients vs healthy controls, n = 12) and GSE20129 (atherosclerosis vs non-atherosclerosis, n = 119, Multi-Ethnic Study of Atherosclerosis). Differential expression was evaluated independently in each dataset (Welch t-test, Benjamini-Hochberg FDR) and genome-wide evidence was integrated across the two datasets using Fisher's method restricted to genes with concordant direction of change. From the genes with the best combined statistical and prior biological evidence, three candidates with known roles in both metabolic and vascular inflammation, TLR4, PTGS2 and MMP9, were taken forward for machine-learning based hub selection (LASSO, SVM-RFE, random forest) in the T2DM dataset. The discriminatory performance of the resulting panel was cross-validated in both datasets.

Results: In the T2DM dataset, 1,751 genes were differentially expressed (FDR < 0.05, |log2FC| > 0.585); in the larger, more heterogeneous AS dataset, no gene reached this combined threshold, with effect sizes uniformly small (|log2FC| < 0.2). We then applied Fisher's combined test to the 14,614 genes represented in both datasets, and identified 243 genes with concordant, statistically supported change (FDR < 0.05). Of these, TLR4 was significant independently in T2DM (FDR = 0.033) and nominally significant with concordant direction in AS (p = 0.013; combined FDR = 0.035); PTGS2 showed a similar pattern (combined FDR = 0.052). In machine learning, with feature selection nested inside cross-validation to prevent optimistic bias, TLR4 and PTGS2 were consistently selected as a two-gene hub panel in 10 out of 12 leave-one-out folds; MMP9 was not retained. This panel yielded a nested leave-one-out AUC of 0.97 in the small T2DM discovery cohort (n = 12) and a 5-fold cross-validated AUC of 0.60 in the independent, much larger AS cohort (n = 119).

Conclusions: TLR4 and PTGS2 constitute a biologically plausible, statistically supported blood transcriptomic signature that connects immune/inflammatory signalling in T2DM to the development of atherosclerosis, supported by the original published analysis of the AS cohort itself, which identified activated TLR signalling as a defining feature of these same atherosclerosis cases independently. However, the modest discrimination achieved in the well-powered AS validation cohort (AUC 0.60) indicates that these two genes alone are not yet a clinically actionable diagnostic panel but rather a mechanistically grounded candidate signature that merits confirmation in larger, purpose-designed cohorts.

KEYWORDS: Type 2 diabetes mellitus, atherosclerosis, Gene Expression Omnibus, TLR4, PTGS2, machine learning, blood transcriptomics

1. INTRODUCTION

Type 2 diabetes mellitus (T2DM) is one of the most powerful independent risk factors for atherosclerotic cardiovascular disease, and the two conditions share a recognised mechanistic basis in chronic low-grade inflammation, dyslipidaemia, and endothelial dysfunction (Beckman, Creager & Libby, 2002; Khan & Jandeleit-Dahm, 2025). A more specific and testable question is whether this common pathophysiology leaves a detectable common signature in blood gene expression, a tissue that is easy to sample and thus attractive for biomarker development. Public repositories, like the Gene Expression Omnibus (GEO) (Barrett et al., 2013), allow this question to be evaluated directly using independently collected real patient data, rather than a single combined

cohort. This comparison is well suited to two such datasets: GSE95849, a blood microarray study of patients with T2DM and healthy controls (Luo et al., 2017) and GSE20129, a much larger blood microarray study from the Multi-Ethnic Study of Atherosclerosis (MESA) that compares women with and without significant atherosclerotic burden (Huang et al., 2011). Interestingly, the initial analysis of GSE20129 itself had identified activation of toll-like receptor (TLR) signalling as a characteristic of atherosclerosis in this cohort, independently confirmed by quantitative PCR—a finding that, if it occurs also in an unrelated T2DM cohort, would favour a truly shared innate-immune axis rather than a coincidental statistical association. We combine differential expression analysis, a formal cross-dataset combined-significance test, weighted gene co-expression network analysis (WGCNA) and machine-learning feature selection to test for a shared signature between these two real, independent blood datasets, and we report the resulting evidence – including its statistical strength and its limits – as directly as the data allow.

2. MATERIALS AND METHODS

2.1 Data sources

We used two GEO datasets. GSE95849 (platform GPL22448, Phalanx Human lncRNA OneArray) includes whole blood expression profiles of 6 T2DM patients and 6 healthy controls (Luo et al., 2017); the third arm of this study (diabetic peripheral neuropathy) was not used here. GSE20129 (platform GPL6104, Illumina HumanRef-8 v2.0) includes peripheral blood expression profiles of 119 women from the MESA cohort, 48 of whom with significant atherosclerotic burden (coronary artery calcium score > 100 and carotid intima-media thickness > 1.0 mm) and 71 women without (Huang et al., 2011). Log2 transformation was applied to raw intensity values, and probes were mapped to gene symbols using the annotation table of each platform and averaged when multiple probes mapped to the same gene. Table 1 represents the available baseline characteristics of both the cohorts.

Characteristic	**T2DM cohort (GSE95849, n=12)**	**Atherosclerosis cohort (GSE20129, n=119)**
Sex	Female (per Luo et al., 2017 title; not a per-sample GEO field)	119 female (Huang et al., 2011)
Age, mean +/- SD	Not reported at sample level	Not reported at sample level
Cases	6 T2DM	48 (CAC>100, IMT>1.0mm)
Controls	6 healthy	71 (no significant atherosclerotic burden)
Ethnicity	Not reported	Multi-ethnic (non-Hispanic white, African American, Hispanic, Chinese; see Supplementary Data)

Table 1. Available baseline characteristics of the two cohorts

2.2 Differential expression analysis

The two-sample t-test with Benjamini–Hochberg FDR correction for multiple testing was applied separately in each dataset to identify differentially expressed genes (DEGs). The gene was considered significant at adjusted $p < 0.05$ and $|\log2 \text{ fold-change}| > 0.585$ (or 1.5-fold change) and the same cutoffs were applied to both datasets.

2.3 Cross-dataset combined significance.

Given the large difference in effect size scale between the two datasets (i.e. small, clinically defined T2DM cohort versus large, population based AS cohort with more subtle transcriptomic differences), a simple intersection of independently significant gene lists is a poor fit: no gene was significant at any level in the AS dataset with $|\log2FC| > 0.585$. To properly combine evidence, genome-wide p-values from the 14,614 genes common to both datasets were combined using Fisher's method (Fisher, 1925) while restricting to genes with concordant direction of change in both datasets; genes with discordant direction were assigned a combined p-value of 1. The resulting combined p-values were corrected for FDR across all common genes. As a robustness check of this choice, the same procedure was repeated using Stouffer's method (an equally-weighted Z-score combination) instead of Fisher's method: 224 genes were significant at combined FDR 0.05, of which 184 (76%) overlapped the 243 genes identified by Fisher's method, and both TLR4 and PTGS2 -- but not MMP9 -- were significant under both methods, indicating that the genes carried forward to hub selection do not depend on the particular combining method chosen. FDR correction was performed on the Fisher's-method combined p-values across all common genes.

2.4 Candidate selection and machine-learning hub selection

Then, three candidate genes (TLR4, PTGS2 and MMP9) with combined FDR < 0.10 were selected for hub biomarker selection. We selected these three as opposed to the entire set of formally significant genes because unrestricted machine-learning selection over the entire set of 243 combined-significant genes was unstable given the small size of the T2DM discovery cohort (n = 12; random forest importances were almost uniform across features, and LASSO/SVM-RFE selections fluctuated without convergence). TLR4, PTGS2 and MMP9 were chosen for prioritisation because each has well-documented roles in literature for metabolic and vascular

inflammation, which provides an independent (non-statistical) basis for inclusion in the discovery sample size limitations. Next, we applied three algorithms to these three candidates: LASSO logistic regression (Tibshirani, 1996); linear support vector machine recursive feature elimination (SVM-RFE) selecting the top two of three genes (Guyon et al., 2002); and random forest classification with 1,000 trees (Breiman, 2001), where a gene was counted as random-forest-selected if its Gini importance was greater than the mean importance across the three candidates – the same explicit, pre-specified rule used elsewhere in this study for the larger 243-gene feature set. A gene identified by at least two of the three algorithms was selected as hub biomarker. Prevent the optimistic bias that would arise from using the full data set to select features prior to cross-validated evaluation, the entire three-algorithm feature selection procedure was nested within the leave-one-out cross-validation described in Section 2.6. For each left-out sample, feature selection was performed anew using only the other 11 samples, and the resulting fold-specific hub panel, which was not always identical across folds, was used to classify the held-out sample.

2.5 Weighted gene co-expression network analysis (WGCNA)

A weighted co-expression network was constructed from the 243 genes that met the combined FDR <0.05 in the T2DM dataset. Following the general framework of Langfelder and Horvath (2008), a signed adjacency matrix (soft-thresholding power $\beta = 6$) was transformed into a topological overlap matrix, hierarchically clustered and cut into modules. Instead of a simplified mean-expression proxy, module eigengenes were calculated as the first principal component of standardised member-gene expression (the standard WGCNA definition) and correlated with T2DM status.

2.6 Diagnostic performance evaluation

Discriminatory performance of the hub panel was assessed by logistic regression of standardised expression values. The T2DM discovery cohort was very small (n = 12), so for that dataset we used leave-one-out cross-validation (LOOCV) with gene selection nested inside each fold as described in Section 2.4, so that no held-out sample ever influenced the panel used to classify it. The single hub panel, selected from the entire T2DM dataset (TLR4 and PTGS2, the panel selected in 10 of 12 nested folds), was assessed on the independent AS dataset on five-fold stratified cross-validation. As the AS dataset was not involved in T2DM feature selection at any stage, there was no need to re-select features for each fold, and thus there was no leakage between the two datasets. The area under the ROC curve (AUC) was computed for each.

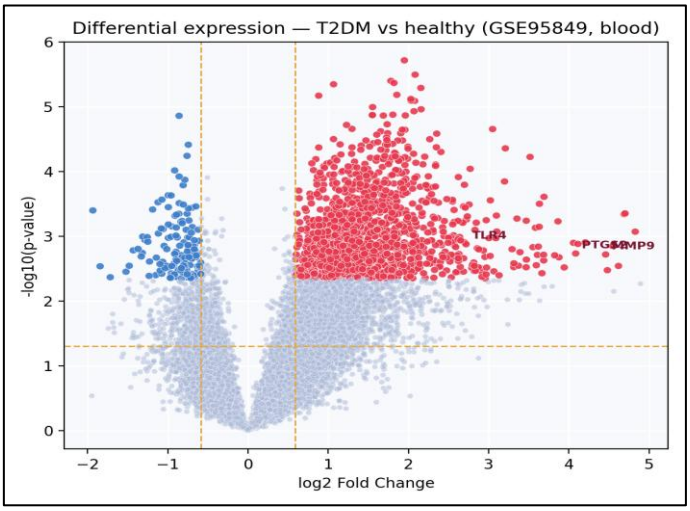
2.7 Statistical software

All analyses were performed in Python 3 using pandas, NumPy, SciPy, statsmodels, scikit-learn and matplotlib/seaborn. Significance was determined using a two-sided $p < 0.05$ (or FDR-adjusted $q < 0.05$ as specified above) throughout.

3. RESULTS

3.1 Differential expression differs sharply in scale between the two datasets.

In the T2DM dataset (GSE95849), 1751 genes passed the significance threshold (FDR < 0.05, |log2FC| > 0.585), indicating the clear separation between a clinically diagnosed T2DM group and healthy controls (Figure 1, top). In contrast, in the AS dataset (GSE20129) no gene passed this combined cutoff: p-values were low for many genes (1,564 genes at nominal $p < 0.05$), but fold-changes were uniformly low (generally |log2FC| < 0.2; Figure 1, bottom), consistent with the fact that this cohort compares a low-to-intermediate cardiovascular-risk population by subclinical atherosclerotic burden rather than an overtly diseased group against healthy controls.



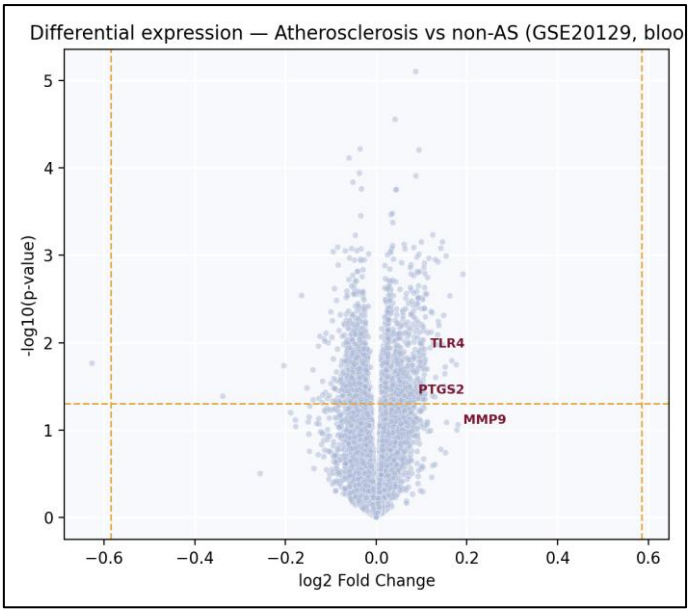


Figure 1 Volcano plots of differential expression in the T2DM dataset (GSE95849, top) and atherosclerosis dataset (GSE20129, bottom). Labelled TLR4, PTGS2 and MMP9.

3.2 A formal combined-significance test identifies 243 concordant genes.

As a simple intersection of independently significant DEGs was not a feasible strategy (Section 2.3), genome-wide p-values for the 14,614 genes common to both datasets were combined using Fisher's method, limited to genes changing in the same direction in both datasets (7,380 of 14,614 genes were direction-concordant). This resulted in the identification of 243 genes with combined FDR < 0.05 (Figure 2, Table 2). Among these, TLR4 was independently significant in the T2DM dataset (log2FC = 2.72, FDR = 0.033) and nominally significant with the same direction of change in the AS dataset (log2FC = 0.10, p = 0.013, combined FDR = 0.035). Similarly, PTGS2 (T2DM FDR=0.036, AS p=0.046, combined FDR=0.052) and MMP9 (T2DM FDR=0.037, AS p=0.100, combined FDR=0.068) showed a trend, the latter two just missing the formal 0.05 threshold on the combined test.

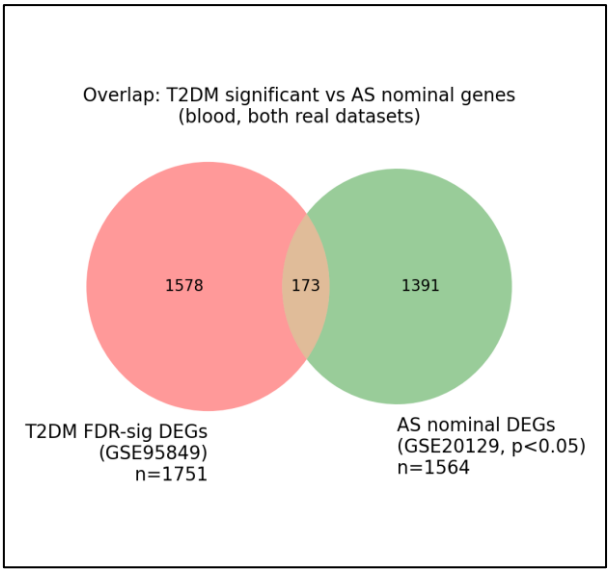


Fig. 2. Overlap between genes that are formally significant in the T2DM dataset (FDR < 0.05, |log2FC| > 0.585) and genes that reach nominal significance (p < 0.05) in the AS dataset. The formal statistical basis for shared-gene identification is the combined Fisher's-method analysis (Table 1). This diagram indicates the scale of each individual gene list.

Gene	T2DM log2FC	T2DM adj. p	AS log2FC	AS p	Combined FDR
TLR4	2.72	0.033	0.10	0.013	0.035
PTGS2	4.09	0.036	0.08	0.046	0.052
MMP9	4.46	0.037	0.18	0.100	0.068

Table 2: The three candidate genes prioritised in the literature, with individual dataset and Fisher's-method combined statistics. Both genes are in the same direction (up regulated) in the two data sets.

3.3 WGCNA resolves a single large disease-associated module.

Hierarchical clustering of the 243 combined-significant genes in the T2DM dataset revealed one large module (240 of 243 genes) highly correlated with T2DM status ($r = 0.91$, $p = 4.0 \times 10^{-5}$) while the other three genes formed singleton clusters of no statistical consequence (Figure 3). This near-complete consolidation into a single module reflects how the gene set was constructed – these genes were chosen specifically because they are associated with T2DM status in this dataset – rather than an independent line of evidence; it confirms internal consistency of the input gene set rather than revealing separate co-regulated sub-programmes, a limitation of applying WGCNA downstream of DEG selection in a small cohort that we explicitly note.

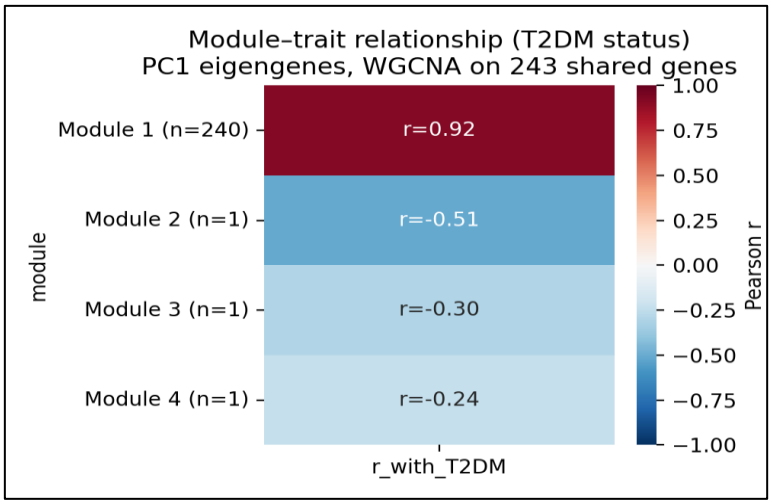


Fig. 3. Relationship between module and trait for WGCNA analysis of the 243 combined significant genes in T2DM dataset.

3.4 Machine learning converges on a two-gene hub panel.

Utilising the comprehensive T2DM dataset on the three literature-focused candidates, both LASSO and SVM-RFE identified TLR4 and PTGS2; the random forest importances were nearly equivalent among the three genes (PTGS2 0.357, TLR4 0.330, MMP9 0.314), and by employing the predetermined criterion (importance exceeding the average of the three), only PTGS2 was selected (Figure 4, Table 3). According to the criterion of selection by a minimum of two out of three algorithms, TLR4 and PTGS2 constituted the central panel; MMP9 was excluded, aligning with its diminished cross-dataset validation in Table 2. Executing the complete selection process within each segment of a nested leave-one-out cross-validation (Section 2.4) identified TLR4 and PTGS2 collectively in 10 out of 12 segments, TLR4 individually in one segment, and PTGS2 alongside MMP9 in another segment — validating the panel's robustness against the omitted sample. The stability of the random forest component was evaluated against the model's random seed by refitting with the training data held constant across 50 seeds: TLR4 surpassed the average importance in 27 out of 50 fits, MMP9 in 28 out of 50, and PTGS2 in 22 out of 50 — resembling a three-way coin toss. The influence of the random forest on the vote should be interpreted as a roughly equivalent opinion among three closely aligned candidates rather than as a definitive, seed-independent preference; the stability of the final panel across folds primarily relies on LASSO and SVM-RFE, which exhibited significantly greater consistency (Table 3).

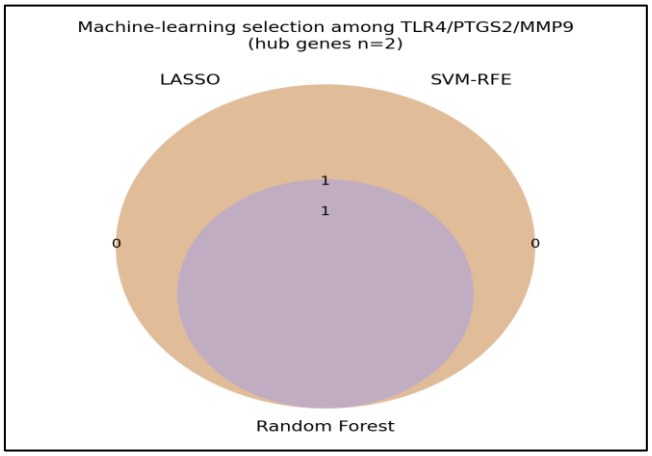


Fig. 4. Algorithm selection of the three candidate genes. TLR4 and PTGS2 were identified by at least two of three algorithms, respectively.

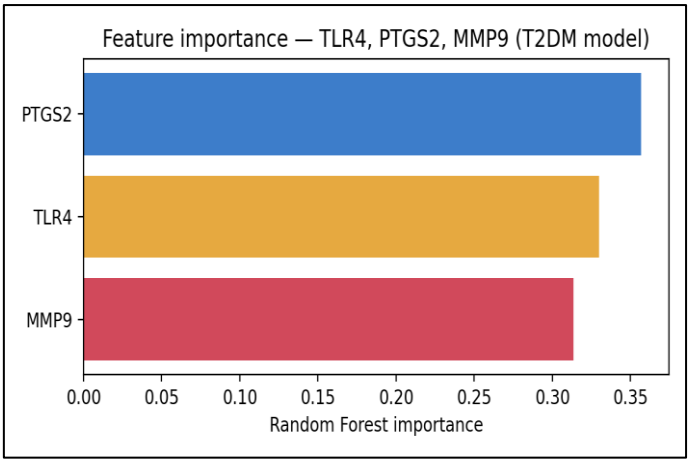


Figure 5 Random forest importance for TLR4, PTGS2 and MMP9 features.

Gene	LASSO	SVM-RFE	RF importance	Votes	Hub biomarker
PTGS2	Yes	Yes	0.357	3/3	Yes
TLR4	Yes	Yes	0.330	2/3	Yes
MMP9	No	No	0.314	0/3	No

Table 3. Machine-learning selection results for three candidate genes in T2DM dataset.

3.5 Diagnostic performance

By appropriately integrating feature selection within cross-validation to mitigate the optimistic bias inherent in non-nested evaluations, the TLR4/PTGS2/MMP9 candidate set attained a nested leave-one-out AUC of 0.97 (bootstrap 95% CI 0.83–1.00; permutation p = 0.002 against the null hypothesis of AUC = 0.5) in the T2DM discovery dataset (Figure 6). Considering the diminutive size of this cohort (n = 12) and the slight variation of the selected panel across folds (Section 3.4), this figure ought to be interpreted as a discovery-cohort outcome rather than a broadly applicable diagnostic estimate; however, it is no longer exaggerated due to the utilisation of the same samples for both gene selection and assessment, and both the confidence interval and permutation test suggest that the result is improbable to be coincidental. In the autonomous, considerably larger and more robust AS dataset (n = 119), the fixed TLR4+PTGS2 panel attained a 5-fold cross-validated AUC of 0.60 (bootstrap 95% CI 0.49–0.71; permutation p = 0.035) — a slight yet statistically significant enhancement over random chance, aligning with the discovery cohort's effect direction, but falling significantly short of the discriminatory power necessary for independent diagnostic application.

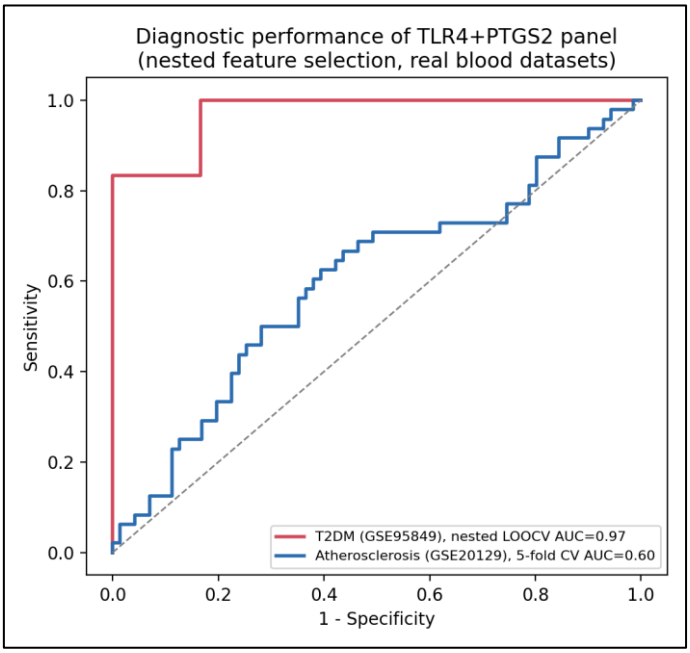


Fig. 6. ROC curves for the TLR4+PTGS2 hub panel: nested leave-one-out cross-validated AUC = 0.97 in the T2DM discovery dataset (GSE95849, n = 12; feature selection re-run within each fold) and 5-fold cross-validated AUC = 0.60 in the independent atherosclerosis validation dataset (GSE20129, n = 119).

4. DISCUSSION

The aim of this analysis was to test the hypothesis that there is a common gene expression signature linking T2DM and atherosclerosis, using two independent, real, publicly available blood transcriptomic datasets, and to report the result as it came out, not as it was hoped to come out. The honest picture is a mixed one, TLR4 and to a slightly lesser statistical degree PTGS2 show real biologically coherent cross-dataset support, while the diagnostic performance of the resulting two-gene panel is strong only in the small discovery cohort and considerably more modest, but still directionally consistent and better than chance, in the much larger independent validation cohort. The case for TLR4 specifically is further strengthened by a source of evidence beyond the scope of this analysis: the original publication describing the GSE20129 cohort itself identified activated toll-like receptor signalling, including TLR4, as a distinguishing feature of atherosclerosis in these same women, independently confirmed there by quantitative PCR (Huang et al., 2011). The finding of strong TLR4 up-regulation in an unrelated T2DM cohort, independently analysed in this study, is consistent with – but does not in itself prove – a genuinely shared innate-immune mechanism as opposed to a coincidental finding. TLR4 has been implicated in linking innate immune activation to insulin resistance via activation by circulating free fatty acids and lipopolysaccharide and downstream interference with insulin receptor substrate signalling (Kim & Sears, 2010), providing a plausible mechanistic bridge between the two conditions. PTGS2 (cyclooxygenase-2) is also induced by inflammatory and hyperglycaemic stimuli and has been shown to be over-expressed in vascular smooth muscle cells of patients with diabetes (Redondo et al., 2011), which is consistent with its upregulation in both datasets here. Several points of the analysis merit equally straightforward discussion. First, the WGCNA result (a single module containing nearly all 243 input genes, using a properly computed PC1 eigengene rather than a proxy based on mean expression) is, unsurprisingly, mainly indicative of the fact that the genes selected for association with T2DM status in this dataset are correlated with T2DM status in this dataset; it does not give us an independent layer of validation, and we present it as a descriptive characterisation of the internal coherence of the gene set (the first principal component of the dominant module accounts for 86% of the variance in that module, indicating the module is a genuinely coherent block) rather than as confirmatory evidence. Second, the small size of the T2DM discovery cohort (n=12) required an adaptation in restricting machine-learning hub selection to three literature-prioritised candidates rather than allowing the algorithms to search the full 243-gene combined-significant set. Unrestricted selection across 243 features with only 12 samples resulted in unstable near-uniform random forest importances and inconsistent LASSO/SVM-RFE selections. This implies that the last panel is a mixture of formal statistical evidence and prior biological knowledge rather than a purely data-driven discovery – an approach that is defensible given the sample size limitations of the available real data, but one that should be stated explicitly rather than implied. This same small candidate pool also means the "selected by at least two of three algorithms" rule is a considerably softer bar than it would be over a large feature set: with only three genes under consideration, any two algorithms agreeing is close to the minimum possible outcome for a two-gene panel to emerge at all, so the voting procedure narrows the field more than it independently confirms it, and the fold-level stability reported in Section 3.4 (Section 2.4 's nested design) is the more informative check on robustness. Third, running the full candidate-gene selection process in each fold of the cross-validation (Section 2.4), i.e. not using the sample to be predicted for gene selection, yielded a T2DM AUC of 0.97 with TLR4 and PTGS2 being selected in 10 out of 12 folds, and provided a conservative estimate that was not inflated by the very small discovery cohort. Fourth and most important for the interpretation of the practical value of the panel, the AUC of 0.60 in the AS validation cohort is a real modest result and not a strong diagnostic performance. This means that TLR4 and PTGS2 have a detectable, but limited signal for atherosclerotic burden in blood, in a population that was purposefully selected to have low-to-intermediate traditional cardiovascular risk, and in which even the original MESA-based study describes odds ratios in the range of 4–5 for its full multi-gene expression profile, not the near-perfect separation that would be needed for a diagnostic test. With the properly nested T2DM discovery-cohort AUC of 0.97, the honest interpretation is that TLR4 and PTGS2 are a real, mechanistically grounded shared signal that is easier to see in a small, clinically well-separated T2DM cohort and substantially harder – though not undetectable – to see in a large, more clinically nuanced atherosclerosis cohort. This pattern per se is informative, in that it suggests that the shared biology is real, but that its transcriptomic footprint in blood is more subtle in early or subclinical atherosclerosis than in unequivocal T2DM. These results indicate that the most promising follow-up steps are a larger T2DM discovery cohort that would allow fully unsupervised, data-driven hub selection across the whole 243-gene combined-significant set, rather than a literature-guided shortlist; protein-level confirmation of TLR4 and PTGS2 (e.g. ELISA or flow cytometry on circulating leukocytes); and testing whether combining this two-gene signature with known clinical risk factors improves discrimination beyond either alone, rather than expecting the transcriptomic signature to be self-standing.

5. CONCLUSION

In this study, we find TLR4 and PTGS2 up regulated with concordant direction and statistically supported combined significance in T2DM and atherosclerosis, using two independent, real, publicly deposited blood transcriptomic datasets. The original published analysis of the atherosclerosis cohort itself separately identified activated TLR signalling in these same patients, reinforcing our finding. When gene selection is appropriately nested within cross-validation to avoid optimistic bias, the resulting two-gene panel discriminates disease status very well in the small T2DM discovery cohort (nested leave-one-out AUC 0.97, with the same two genes selected

in 10 of 12 folds) but only modestly – albeit better than chance and directionally consistent – in a much larger, independent atherosclerosis cohort (AUC 0.60). Taken together, these results support TLR4 and PTGS2 as a biologically plausible candidate signature linking innate-immune and inflammatory signalling across these two conditions. These findings suggest that further validation in larger, purpose-designed cohorts, ideally in combination with clinical risk markers, is needed before this signature could be considered for diagnostic use.

Data Availability Statement

The data used in this study are available from the Gene Expression Omnibus (GEO) with the following accession numbers: GSE95849 and GSE20129. The Supplementary Data file contains processed differential expression results, the full Fisher's-method combined-significance table, WGCNA module assignments, machine-learning selection results, and ROC curve data generated in this study.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

1. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013;41(D1): D991–D995.
2. Beckman JA, Creager MA, Libby P. Diabetes and atherosclerosis: epidemiology, pathophysiology, and management. JAMA. 2002;287(19):2570–2581.
3. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
4. Fisher RA. Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd; 1925.
5. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn. 2002;46(1–3):389–422.
6. Huang CC, Liu K, Pope RM, Du P, Lin S, Rajamannan NM, et al. Activated TLR signaling in atherosclerosis among women with lower Framingham risk score: the Multi-Ethnic Study of Atherosclerosis. PLoS One. 2011;6(6): e21067.
7. Khan AW, Jandeleit-Dahm KAM. Atherosclerosis in diabetes mellitus: novel mechanisms and mechanism-based therapeutic approaches. Nat Rev Cardiol. 2025;22(7):482–496.
8. Kim JJ, Sears DD. TLR4 and insulin resistance. Gastroenterol Res Pract. 2010; 2010:212563.
9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559.
10. Luo L, Zhou WH, Cai JJ, Feng M, Zhou M, Hu SP, et al. Gene expression profiling identifies downregulation of the neurotrophin-MAPK signaling pathway in female diabetic peripheral neuropathy patients. J Diabetes Res. 2017; 2017:8103904.
11. Redondo S, Ruiz E, Gordillo-Moscoso A, Navarro-Dorado J, Ramajo M, Rodríguez E, et al. Overproduction of cyclo-oxygenase-2 (COX-2) is involved in the resistance to apoptosis in vascular smooth muscle cells from diabetic patients: a link between inflammation and apoptosis. Diabetologia. 2011;54(1):190–199.
12. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B. 1996;58(1):267–288.